



**XXIII SNPTEE
SEMINÁRIO NACIONAL
DE PRODUÇÃO E
TRANSMISSÃO DE
ENERGIA ELÉTRICA**

FI/GGH/16
18 a 21 de Outubro de 2015
Foz do Iguaçu - PR

GRUPO - I

GRUPO DE ESTUDO DE GERAÇÃO HIDRÁULICA- GGH

**USO DE TÉCNICAS DE MINERAÇÃO DE DADOS PARA EXTRAÇÃO DE INDICAÇÕES DE FALHA NA
OPERAÇÃO DE HIDROGERADORES A PARTIR DE MEDIDAS DE DESCARGAS PARCIAIS**

**Ana Carolina Neves Pardauil(*)
UFPA**

**Ubiratan Holanda Bezerra
UFPA**

**Werboston Douglas de Oliveira
UFPA**

**Hélio de Paiva Amorim Jr
CEPEL**

**André Tomaz de Carvalho
CEPEL**

RESUMO

Medição e análise de descargas parciais é um dos métodos mais eficazes para análise do isolamento estatórico do gerador. No entanto, apesar de possuírem padrões bem definidos, não é trivial encaixar os sinais nestes padrões, devido principalmente ao grande número e variedades de ocorrências de DPs. O uso de uma ferramenta que agilize este processo é proposto neste trabalho, baseado em técnicas de mineração de dados, que é uma solução para análise de grandes volumes de informações. Foi utilizada uma abordagem híbrida (não-supervisionado/supervisionado) para identificar padrões e posteriormente classifica-los dentre as formas conhecidas de DPs.

PALAVRAS-CHAVE

Descarga Parcial, UHE Tucuruí, Mineração de Dados, Clusterização, Árvore de Decisão.

1.0 - INTRODUÇÃO

A maior causa de falhas em enrolamentos estatóricos de hidrogeradores está relacionada com a sua isolamento elétrica. Assim é necessário procurar estabelecer meios para que haja um monitoramento seguro e um diagnóstico eficiente para que essas falhas sejam detectadas ainda em fase incipiente, de modo a evitar danos irreversíveis ao equipamento, com a consequente interrupção da operação.

A medição e análise das descargas parciais (DP) é um dos métodos mais seguros, eficazes e confiáveis de se fazer o monitoramento e diagnóstico do estado do enrolamento (1). Com a grande pressão do mercado para que haja um maior fornecimento de energia, têm-se a necessidade de se utilizar geradores ao máximo de sua capacidade, o que vem ocasionando maior solicitação dessas máquinas e consequentemente estas ficam mais suscetíveis à falhas.

Considerando este cenário, este artigo tem por objetivo principal estabelecer métodos de diagnóstico de falhas incipientes nos enrolamentos do estator, com base nos tipos e magnitudes das descargas parciais, utilizando técnicas de mineração de dados. No desenvolvimento dos procedimentos aqui propostos, serão utilizadas as normas e os padrões de descarga já pré-determinados, bem como as medições realizadas nas máquinas das casas de força I e II da Usina Hidrelétrica de Tucuruí, as quais foram obtidas pelo sistema do IMA-DP de medição de descargas parciais, sistema este desenvolvido pelo CEPEL.

Apesar de o método de medição das descargas parciais ser consolidado, os diagnósticos apresentados a respeito da isolamento do estator tem sua fundamentação bastante empírica, ou seja, necessita-se de um especialista para analisar todos os gráficos obtidos para determinar em quais padrões cada descarga se encaixaria e qual a gravidade da mesma naquele equipamento, o que consequentemente torna-se algo demorado e sem uma segurança total nas análises apresentadas.

(*) Rua Sergipe, nº 01 – Vila Permanente – CEP 68455-759 Tucuruí, PA, – Brasil
Tel: (+55 94) 98138-4664 Email: ana_pardauil@hotmail.com

Em virtude desta realidade utilizou-se técnicas de mineração de dados, que tem o intuito de criar uma ferramenta inteligente de reconhecimento de padrões de descarga parcial. Após treinadas, as ferramentas baseadas em inteligência computacional poderão ser utilizadas para gerar diagnósticos para qualquer outra máquina com características similares, tornando as análises menos onerosas, mais ágeis e simples. Será apresentada uma abordagem baseada em clusterização e árvores de decisão, afim de agrupar e classificar os padrões de descargas, de maneira ágil e confiável, utilizando dados reais de DPs. Tal proposta se mostrou uma alternativa viável, com bons resultados e de implementação sem grandes dificuldades.

2.0 - DESCARGAS PARCIAIS (DP_s)

As descargas parciais podem ter sua definição como “faíscas” envolvidas em um fluxo de elétrons e íons em pequenos volumes de ar encapsulados em cavidades que se configuram como defeitos no sistema de isolamento que se caracteriza como um espaço não impregnado por material isolante e que pode enclausurar gases provenientes do material isolante(1). Assim, se houver uma cavidade, no interior ou superfície de uma isolação, pode culminar no aparecimento de descargas, pois implicará em diferença de potencial através desta (2,3).

Diante disso normas são estabelecidas e devem ser seguidas para o embasamento dos conceitos e padrões de descargas parciais em máquinas rotativas. Dentre estas se destacam as normas IEC 60270, IEC 60034-27-2 e a IEEE 1434. Ao se tratar dos padrões estabelecidos de descargas parciais que serão de fundamental importância neste trabalho, os quais são apresentados na figura 1, utilizar-se-ão os gráficos provenientes de histogramas de amplitude baseado na taxa de repetição e magnitude da descarga para identificar esses padrões. Tal escolha foi feita para simplificar a leitura e organização dos dados de DPs a serem tratados pelas técnicas de mineração de dados aqui propostas.

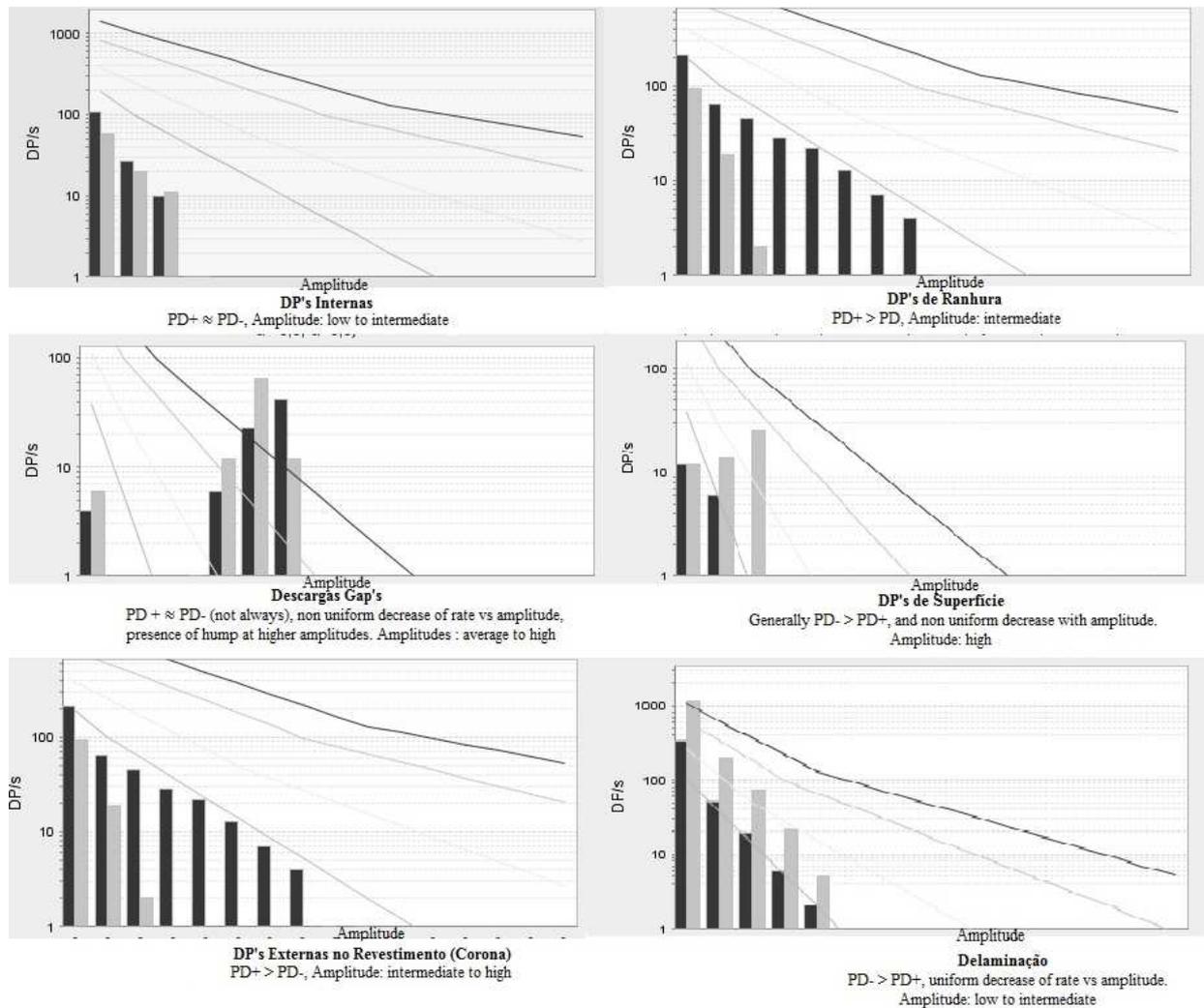


FIGURA 1 - Padrões de Descargas Parciais (4)

Importante frisar que os locais de ocorrência das descargas parciais estão ao longo da barra/bobina estática, tal como exemplificado na figura 2.

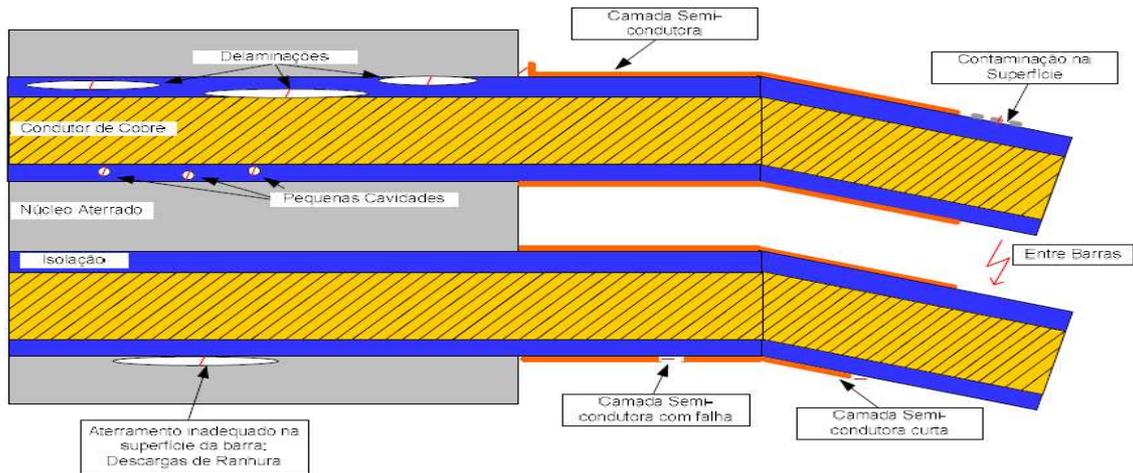


FIGURA 2 - Fontes de Descargas Parciais no Isolamento do Estator (2)

A medição de descargas parciais estabelece as condições de degradação da isolação do enrolamento do estator e diante disso, novos sistemas de medição são criados e/ou aprimorados, afim de ter resultados de medições mais claros e confiáveis. Estas medições são realizadas através de um computador, o qual se conecta através de cabos coaxiais aos terminais da máquina que por sua vez são o caminho para se comunicar com os acopladores capacitivos instalados no estator do hidrogenador que captam o sinal de descarga parcial.

Através da ferramenta de medição do CEPEL, o IMA-DP intensivo, este o qual fica instalado em cada gerador por um período mensal, realiza medições constantes, o que possibilita analisar no tempo em quais ocasiões de operação do gerador as descargas parciais são mais intensas ou brandas. Os resultados das medições são mapas estatísticos, os quais para este trabalho foram transformados em histogramas de amplitude. Ressalta-se que a máquina escolhida apresenta 864 ranhuras, 24 sensores distribuídos 8 por fase. Na figura 3 apresenta-se o arranjo típico de medição empregado, com o sistema IMA-DP, apresentando-se em destaque na tela, o mapa estatístico de magnitude e fase característico da DP em análise.



FIGURA 3 - Instalação e tela do sistema IMA-DP Intensivo, exemplificando o mapa estatístico da DP.

3.0 - MINERAÇÃO DE DADOS

O processo aqui proposto de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados é conhecido como Mineração de Dados (5), ou seja, é uma estratégia eficaz para encontrar informações úteis embutidas em GRANDES volumes de dados. Mineração de dados (MD) é parte integrante de um processo maior, conhecido como Descoberta de Conhecimento em Banco de Dados (KDD, do inglês Knowledge Discovery in Databases), que por sua vez consiste em uma série de passos de transformações de dados (6). A MD torna-se uma técnica bastante interessante pois:

- Frequentemente existe informação "escondida" nos dados que não é evidente de ser encontrada utilizando linguagens de consultas tradicionais, ou seja, podem haver algumas relações até então desconhecidas nas bases de dados.

- Analistas humanos podem levar semanas para correlacionar e descobrir alguma informação útil dentro de uma grande massa de dados.
- Boa parte dos dados nunca é analisado: “cemitério” de dados.

Basicamente, a MD possui dois tipos de tarefas:

- Tarefas Preditivas - prever o valor de um determinado atributo baseado nos valores de outros atributos, por exemplo Classificação supervisionada- árvores de decisão, redes neurais, SVMs.
- Tarefas Descritivas - Derivar « padrões »: correlações, tendências, anomalias, agrupamentos dentro de uma grande massa de dados, como exemplos têm-se: Regras de Associação – Padrões Sequenciais – Agrupamentos - Anomalias

Neste trabalho, será usado uma forma híbrida dessas tarefas, utilizando uma técnica para realizar a tarefa descritiva (Clusterização) com a finalidade de agrupar padrões, e em seguida, usa-se uma árvore de decisão (técnica de classificação) para utilização como ferramenta preditiva.

3.1 Clusterização

Clusterização, ou simplesmente agrupamento, é uma tarefa que tem como finalidade agrupar registros semelhantes criando novas classes. Ela é uma tarefa de mineração muito útil, haja vista, que instintivamente as pessoas visualizam os dados segmentados em grupos discretos, como por exemplo, tipos de carros ou motos. Clusterização é um processo de classificação não supervisionada de padrões em grupos chamados de clusters. Os clusters são formados por dados com características semelhantes, a partir das relações que medem a distância entre um par de padrões no espaço de características, podendo ser medida pela distância euclidiana. Neste trabalho foi utilizado o K-MEANS, onde faz-se os seguintes passos (7,8,9):

1. Escolher k pontos, dentro do espaço de características, representando os centros dos k clusters em que é desejado dividir o conjunto de dados.
2. Assinar cada padrão ao centro mais próximo de acordo com a função de similaridade.
3. Recalcular os centros dos clusters utilizando os dados membros de cada cluster.
4. Repetir o algoritmo desde o item dois até atingir um critério de parada.

3.2 Árvore de Decisão

A árvore de decisão (DT, do inglês Decision Tree) para fins de classificação é uma técnica de aprendizado de máquina supervisionado para solucionar problemas com dados de alta dimensionalidade (10). O princípio fundamental é obter um modelo preditivo para classificar um objetivo usando os atributos que contribuem diretamente para tal objetivo. A DT converte um complexo processo de classificação em algumas instruções lógicas “if-then”, em termos dos limites dos atributos de entrada ou de suas combinações lineares. O modelo da árvore de decisão tem uma estrutura binária com dois tipos de nós, o nó interno com dois sucessores e o nó terminal sem qualquer sucessor. Para cada nó terminal (folha), um resultado de classificação será atribuído em termos da classe majoritária do objetivo. O processo de classificação começa a partir do nó raiz e finaliza em um nó terminal, onde o resultado de classificação é alcançado. Após a criação da DT, um processo de poda é realizado para a retirada de nós desnecessários e, por fim, diminuir o tamanho final da DT.

Existem diversas variações para construção de DTs, e neste trabalho utilizou-se o CART (árvore de classificação), com o cálculo das divisões dos nós sendo realizados pelo índice GINI, sendo que este método visa minimizar a impureza de cada nó (entropia). Onde a impureza do nó é máxima quando todas as classes possuem igual distribuição e mínima quando existe apenas uma classe.

4.0 - METODOLOGIA

A metodologia proposta, consiste em desenvolver uma ferramenta de auxílio na análise de dados de descargas parciais em hidrogeradores. Tal ferramenta consiste de um encadeamento de tarefas de mineração de dados, afim de fornecer padrões e classifica-los dentre os padrões definidos pela norma de maneira automática. Dessa forma, é possível ganhar agilidade no processo de identificação e análise dos problemas encontrados.

A figura 4 apresenta um esquemático da metodologia proposta, na qual os dados das descargas são coletados pelo IMA-DP por meio dos mapas estatísticos. Em seguida, esses arquivos contendo os mapas estáticos são processados para que fiquem na forma de histogramas de amplitudes das descargas parciais (HDP), gerando assim uma base de dados (BDHDP). A partir de uma extensa base de dados, ou seja, um grande conjunto de informações das descargas nos geradores, utiliza-se a técnica K-means para identificar e agrupar os padrões, e assim cada amostra da BDHDP é rotulada como um dos padrões encontrados na clusterização (7 no total). A partir dos dados rotulados, uma DT é treinada afim de classificar cada padrão, a partir de poucas variáveis, facilitando assim o trabalho do analista.

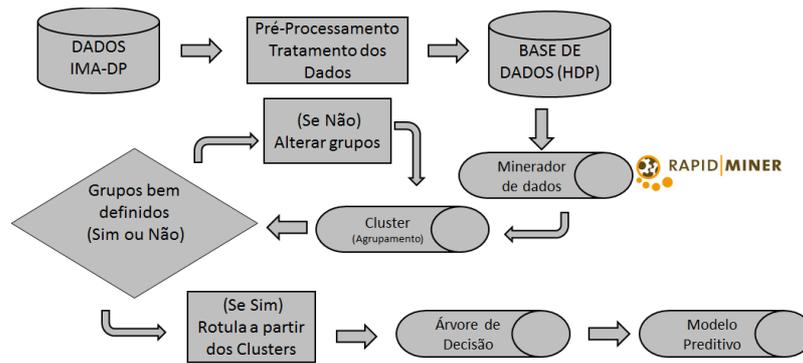


FIGURA 4 - Esquemático da Metodologia Proposta para Identificação de Descargas Parciais.

5.0 - RESULTADOS

Utilizando-se o software *Rapid Miner* aplicou-se a metodologia tal como apresentada na figura 4, foram inseridos 2435 dados de medições os quais representavam informações dos 8 acopladores da fase A de um hidrogerador da Casa de Força 1 da UHE de Tucuruí, destacando-se que o método pode ser replicado as demais fases.

Através do Sistema do IMA-DP, cujo software está instalado em um PXI, este consiste em uma plataforma baseada em PC para sistemas de medição e automação, é captado os sinais provenientes dos acopladores e estes sinais de DPs são traduzidos em mapas estatísticos que apresentam a magnitude por fase conforme ilustrado na figura 3. Em seguida esses dados são processados de mapas estatísticos para histogramas de amplitude em base logarítmica para a classificação dos padrões, estabelecidos pela norma e apresentados na figura 1. Deste modo constitui-se uma base dados consistente e ampla. Através desta gama de dados começa-se o processo de mineração de dados que consiste em averiguar os padrões e separá-los em clusters (grupos), que neste caso foram escolhidos 7 grupos, que representam 6 padrões definidos por norma e 1 padrão que encaixaria casos que possuem níveis baixos de descarga parcial e que não apresentam um padrão definido. É importante salientar que o embasamento e treinamento da ferramenta foi fundamentada em dados reais, ou seja, quando se determinou o número de casos já era esperado que um dos padrões não aparecesse, no caso o padrão de delaminação, porém ao se replicar a técnica para as demais fases do gerador e por conseguinte para as demais máquinas este padrão tenderá a aparecer com a inserção de novos dados ao sistema.

Tendo os clusters definidos chegou-se a casos de padrões de descargas parciais, apresentados na figura 5. Nota-se que os centroides (cada centroide representa um grupo) encontrados pelo algoritmo K-means, representam de forma satisfatória cada um dos padrões de DPs que ocorreram na fase "A" do hidrogerador estudado.

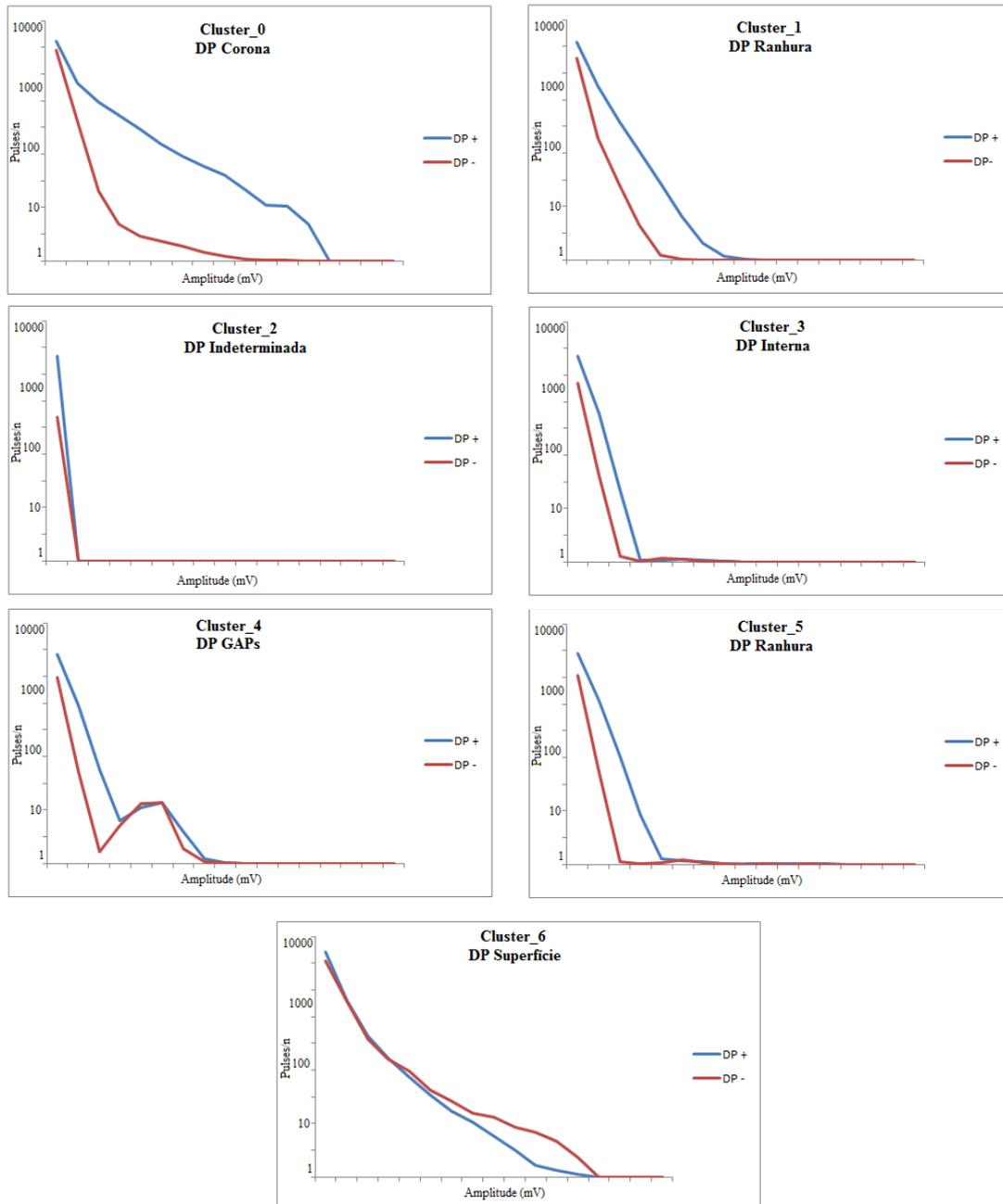


FIGURA 5 – Clusters e Correspondência com Padrões na Norma

A partir da definição dos clusters, pode-se utilizar cada um deles como rotulo para o treinamento de uma máquina de aprendizado supervisionada, tornando-se possível o treino de uma árvore de decisão, que após treinada venha auxiliar a identificação de novas medições em cada padrão determinado, trabalhando com um número menor de variáveis, resultando em uma leitura mais rápida e eficaz. Esses aspectos podem contribuir para melhorar a certeza nos diagnósticos que serão realizados e por conseguinte tem-se um modelo preditivo do sistema, baseado em um conjunto de regras do tipo “se-então”. A figura 6 apresenta um ramo da DT criada para a classificação das DPs, contendo as regras que são seguidas para a classificar os padrões de DPs, isto é, quais os valores no histograma de amplitudes correspondem aos padrões a serem averiguados em cada caso.

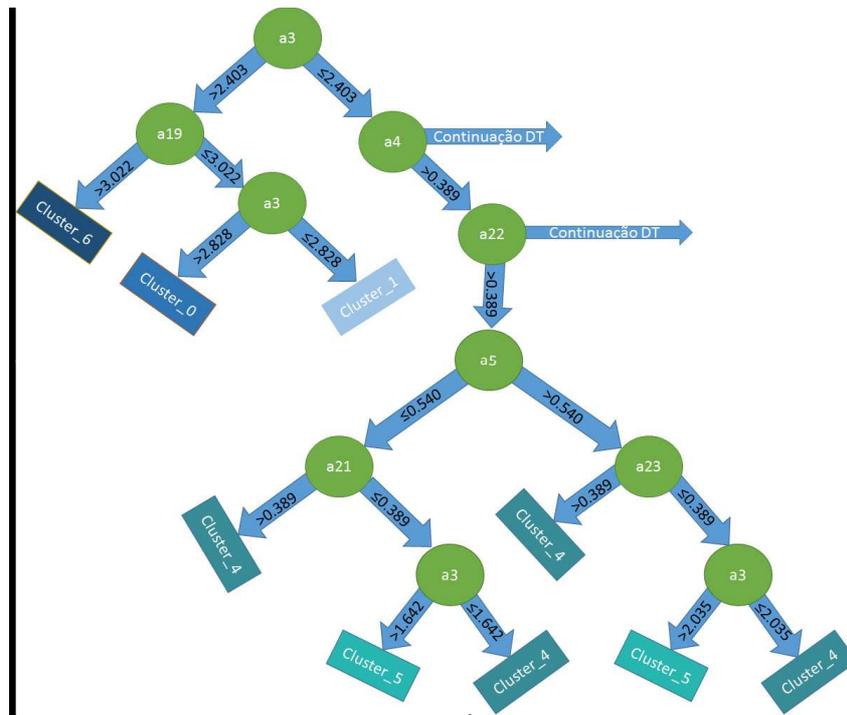


FIGURA 6 – Trecho da Árvore de Decisão

Para o treino, teste e validação da DT, foi utilizada a técnica *K-fold crossvalidation*, onde comprovou-se que o modelo criado pela DT, representa de forma bastante adequada a classificação das DPs. É possível comprovar isto pela matriz de confusão, apresentada na figura 7, a qual mostra a alta taxa de acerto obtida para todos os padrões classificados. A menor taxa de acerto ocorreu para o padrão referente ao cluster 4, sendo 92,42%, tendo-se obtido valores de 100% para a maioria dos padrões analisados, resultando uma taxa de acerto global (accuracy) superior a 98%.

Accuracy: 98,54% +/- 0.75% (mikro:98,54%)								
	True cluster_4	Class precision						
Pred. Cluster_4	183	5	0	9	1	0	0	92,42%
Pred. Cluster_3	5	364	0	4	0	0	0	97,59
Pred. Cluster_0	0	0	379	0	0	0	0	100%
Pred. Cluster_5	10	2	0	607	0	0	0	98,06%
Pred. Cluster_1	0	0	0	0	382	0	0	100,00%
Pred. Cluster_6	0	0	0	0	0	385	0	100,00%
Pred. Cluster_2	0	0	0	0	0	0	135	100,00%
Class Recall	92,42%	98,11%	100,00%	97,90%	99,74%	100%	100,00%	

FIGURA 7 – Matriz de Confusão

6.0 - CONCLUSÃO

Devido à necessidade de se aprimorar os diagnósticos do enrolamento estatórico embasados nas medições de descargas parciais desenvolveu-se uma ferramenta inteligente que possibilitará a identificação de tipos de descargas parciais como apresentados nas normas. Isto facilita um diagnóstico, pois além de averiguar a gravidade de DP alicerçada na sua magnitude pode-se então saber a localização desta na barra e/ou bobina.

Utilizando-se dos dados fornecidos através do Sistema do IMA-DP e fazendo os devidos tratamentos nestes foi possível transformar os mapas estatísticos em histogramas de amplitude, os quais são mais simples e facilitam o

reconhecimento dos padrões. Esta escolha de representar em histogramas também foi feita pelo fato da grande quantidade de dados que seriam originalmente utilizados no treinamento do sistema. Por sinal, o volume de dados foi um dos motivos de se trabalhar apenas com a fase A nesta fase do trabalho, a qual é composta da análise de 8 acopladores, mas ressalta-se que futuramente irá se replicar os estudos para as demais fases da máquina em questão, e posteriormente para as demais máquinas da UHE de Tucuruí.

A utilização de dados reais foi crucial para garantir a veracidade do método, pois nas análises realizadas todos os padrões estavam presentes, e o sistema de classificação aqui proposto os reconheceu e apresentou através dos Clusters todas estas categorias de forma bem definida.

Diante destes resultados animadores pode-se então obter uma árvore de decisão que bem treinada possibilita reconhecer padrões de descargas parciais ao se inserir novos dados ao sistema, tornando assim o trabalho para o analista menos oneroso e mais ágil.

Em trabalhos futuros ampliar-se-á as análises não apenas embasados nos histogramas, mas também nos mapas estatísticos, e espera-se replicar a ferramenta para as demais máquinas da UHE de Tucuruí, bem como inserir na linha de análise, padrões para determinar a gravidade da descarga parcial embasada na magnitude atrelada ao reconhecimento de padrões, pois cada padrão tem um nível de gravidade diferente do outro devido a sua localização no enrolamento, e assim deste modo emitir um diagnóstico mais completo e real da situação do isolamento elétrico do gerador.

7.0 - REFERÊNCIAS BIBLIOGRÁFICAS

- (1) STONE, G.C., "Theory of The Partial Discharge Analyser Test", Presented to EPRI/CEA Workshop on the PDA and RF Monitoring, Toronto, Canada, 1986.
- (2) BRASIL, F.S., "Estudo Experimental de Técnicas de Medição de Descargas Parciais nos Enrolamentos Estatóricos de Hidrogeradores." Tese de Mestrado, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal do Pará, Belém, Brasil, 2013.
- (3) INTERNATIONAL ELETROTECHNICAL COMMISSION, "IEC 60270 High Voltage Test Techniques – Partial Discharges Measurements". Suíça, 2000
- (4) INSTITUTE OF ELECTRICAL AND ELECTRONIC ENGINEERS, "IEEE Std 1434 Guide to Measurement of Partial Discharge in Rotating Machinery". USA, 2014
- (5) FAYYAD, U.M. "Data Mining and Knowledge Discovery in Databases: Applications in Astronomy and Planetary Science". In: Association for Advancement of Artificial Intelligence (AAAI), Redmond, USA, 1996.
- (6) Tan, P-N.; Steinbach, M.; Kumar, V.. "Introdução ao Data Mining." Ciência Moderna, 2009.
- (7) DILLY, R. "Data Mining: an introduction." Belfast: Parallel Computer Centre, Queens University, 1999
- (8) Addrians, P. & Zantinge, D. "Data Mining". Inglaterra: Addison-Wesley., 1996
- (9) Mannila, H. "Data mining: machine learning, statistics and databases." International Conference on Statistics and Scientific Database Management., 1996
- (10) Carpentier, J., "Static Security Assessment and Control: a Short Survey Athens Power Tech", APT 93. Proceedings. Joint International Power Conference, 1993.

8.0 - DADOS BIOGRÁFICOS

Ana Carolina Neves Pardauil nasceu em Macapá-AP no dia 12 de janeiro de 1992, formou-se como Engenheira Eletricista graduada pela UFPA em Tucuruí no ano de 2014, estagiou em 2012 na Eletronorte na UHE de Tucuruí na área de engenharia com ênfase em manutenções de hidrogeradores, medições e análises de descargas parciais. No mesmo ano, após a graduação, iniciou o mestrado de engenharia elétrica pela UFPA na linha de pesquisa de sistemas elétricos de potência. Neste mesmo período ingressou como bolsista do CEPEL realizando medições, pesquisas e análises no segmento de descargas parciais em hidrogeradores.